

Прогнозирование вегетационного индекса растительности сельскохозяйственных угодий Волгоградской области с помощью нейросетевых методов

Е.А. Шурлаева, Д.А. Барышев, А.С. Зубанков, В.Л. Розалиев

Волгоградский государственный технический университет

Аннотация: В работе освещается тема мониторинга состояния растительности с помощью спутниковых технологий. Рассматривается построение вегетационных карт полей NDVI. Для своевременного обнаружения неоднородных и дефектных участков растительного покрова предлагается дополнять спутниковые снимки вегетационными изображениями, сформированными по спрогнозированным значениям индекса растительности. В работе рассматриваются методы прогнозирования индекса NDVI по данным сельскохозяйственных угодий Волгоградской области. Получены результаты обучения рекуррентной нейронной сети с механизмом LSTM, а также результаты обучения алгоритма XGBoost. По итогам обучения определены наиболее важные влияющие на NDVI погодные параметры. Эффективность обученных моделей оценивалась по метрике RMSE.

Ключевые слова: точное земледелие, вегетационные индексы, NDVI, прогнозирование, временной ряд, LSTM, случайный лес.

В настоящее время по всему миру в сельском хозяйстве используются технологии точного земледелия. Точное земледелие (прецизионное, координатное) земледелие – это системный подход к управлению урожайностью, основанный на применении компьютерных и спутниковых технологий с учетом различных условий на разных участках поля. С его помощью можно повысить валовый сбор, сократить затраты, получить качественную продукцию, улучшить физикохимические свойства почвы [1]. Одним из важнейших аспектов точного земледелия является дистанционное зондирование Земли (далее ДЗЗ) с помощью спутниковых технологий. Это обусловлено рядом уникальных особенностей ДЗЗ: широкий территориальный охват, оперативность, контроль за отдаленными районами, всепогодность при использовании радарной съемки и так далее [2]. Кроме снимков, получаемых непосредственно с помощью электронной аппаратуры спутника, для мониторинга за состоянием растительности также

используются индексные изображения, несущие информацию о новых свойствах объектов незаметных на оригинальных снимках. Всего существует более 250 вегетационных индексов, нацеленных на выделение того или иного признака [3]. Но чаще всего для построения вегетационных изображений используют значения нормализованного вегетационного индекса (Normalized Difference Vegetation Index, далее – NDVI). Эта популярность и широкое использование связаны с тем, что NDVI можно рассчитать с помощью любого мультиспектрального датчика с видимым и ближним ИК-диапазоном [4]. Также плюсом NDVI является чувствительность по отношению к зеленой растительности, даже для участков с низкой плотностью растительного покрова [5].

NDVI рассчитывается по формуле:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

где,

- NIR - отражение в ближней инфракрасной области спектра
- RED - отражение в красной области спектра.

Значения NDVI варьируются от – 1 до 1, однако отрицательные значения данный индекс принимает для водоемов, каменистых, или бетонных поверхностей, песка и т.д. Значения NDVI для растительности находятся в диапазоне от 0 до 1.

Пример вегетационного изображения NDVI представлен на рисунке 1.

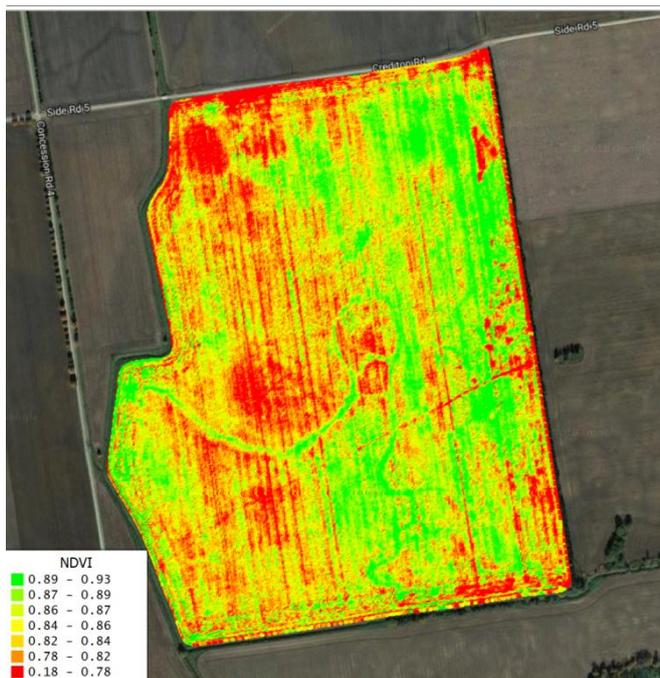


Рис. 1. – Вегетационное изображение NDVI и легенда цветовых значений NDVI.

Подобные вегетационные изображения помогают определить качество растительного покрова, выявить дефектные (неоднородные) участки поля и оценить примерную урожайность поля. Как было сказано ранее, вегетационные изображения NDVI формируют по спутниковым снимкам. Частота получения снимков зависит от частоты прохода спутника над наблюдаемой точкой Земли. Перерыв между снимками может длиться от 3 дней до недели, если рассматривать различные миссии спутников, такие как Landsat 8 или Sentinel 2. Следовательно, получаемые изображения отражают уже прошлое состояние растительности, или в лучшем случае - ее состояние на текущий момент. Но для фермерских хозяйств необходимо получать не только самую актуальную информацию о состоянии растительности, но и прогноз дальнейшего развития растительного покрова. Решить данную проблему возможно с помощью прогнозирования значений NDVI на каждый пиксель спутникового снимка и дальнейшего построения вегетационного

изображения по спрогнозированным значениям индекса растительности. В данной работе рассматриваются методы прогнозирования NDVI для определения будущего состояния растительности на примере Волгоградской области.

NDVI представляет собой временной ряд [6-8]. Следовательно, необходимо рассмотреть общие методы прогнозирования временных рядов и их применение для расчета NDVI значений. Основными статистическими методами исследования временных рядов являются: метод выделения тренда (временного сглаживания), регрессионный, автокорреляционный, адаптивный (скользящих средних), метод гармонического анализа, сингулярного спектрального анализа, бутстрепа (численного размножения выборок) и нейросетевой [9]. В [10-12] рассматриваются классические статистические методы прогнозирования временных рядов для прогнозирования NDVI, такие, как: регрессионный метод, метод временного сглаживания, также комбинации этих методов в математических моделях.

Однако лучшие результаты показывают нейросетевые методы машинного обучения, такие, как: рекуррентные нейронные сети с применением Long short-term memory (долгой краткосрочной памяти, далее – LSTM), и различные реализации алгоритма случайного леса (Random Forest) [13-15].

Для прогнозирования значений NDVI, в данной работе также были выбраны методы машинного обучения: улучшенный алгоритм случайного леса – Extreme Gradient Boosting (метод экстремального бустинга, далее – XGBoost), а также рекуррентные нейронные сети с использованием LSTM.

Для начала необходимо определиться с параметрами, влияющими на состояние растительности, и как следствие, на значения NDVI. В аналогичных исследованиях [13-15] выделяют погодные параметры, такие, как: температура, скорость ветра, давление, влажность и другие. В данной

работе также было принято решение рассмотреть погодные параметры в качестве признаков, подаваемых на вход обучаемых моделей.

Исторические данные о погодных условиях были взяты с worldweatheronline.com. Временной интервал составил 5 лет: с 2015 по 2020 годы, а областью исследования являлась Волгоградская область. Сформированный датасет содержит следующие метеорологические данные:

- максимальная температура за текущие сутки;
- минимальная температура за текущие сутки;
- общее количество выпавшего снега в см;
- общий солнечный час;
- ультрафиолетовый индекс;
- лунное освещение (в %);
- температура точки росы в градусах Цельсия;
- «ощущается как» (температура в градусах Цельсия);
- индекс тепла, измеряемый в градусах Цельсия;
- температура холодного ветра в градусах Цельсия;
- порыв ветра в километрах в час;
- величина облачного покрова в процентах;
- влажность в процентах;
- осадки в миллиметрах;
- атмосферное давление в миллибарах;
- температура в градусах Цельсия;
- видимость в километрах;
- направление ветра в градусах;
- скорость ветра в километрах в час.

Кроме погодных параметров, в качестве входных признаков были выбраны непосредственно значения NDVI на каждый пиксель спутникового

снимка, полученного за прошлый период для учета последовательных изменений в растительном покрове. В качестве датасета вегетационных изображений NDVI используются данные с портала apps.sentinel-hub.com, полученные за 2015-2020 годы по Волгоградской области.

Рассмотрим обучение с помощью алгоритма XGBoost. XGBoost – это улучшенный алгоритм градиентного бустинга (Gradient Boosting). Оба этих алгоритма являются разновидностями алгоритма Random Forest (Случайный лес). Случайные леса могут использоваться не только для задач классификации и регрессии, но и для задач выявления наиболее информативных признаков, кластеризации, выделения аномальных наблюдений и определения прототипов классов [16]. Благодаря функции назначения цифровых весов для каждого признака, можно выделить наиболее важные, влияющие на результат погодные условия.

В результате обучения были определены наиболее важные погодные параметры. На рисунке 2 представлена гистограмма распределения важности каждого признака.

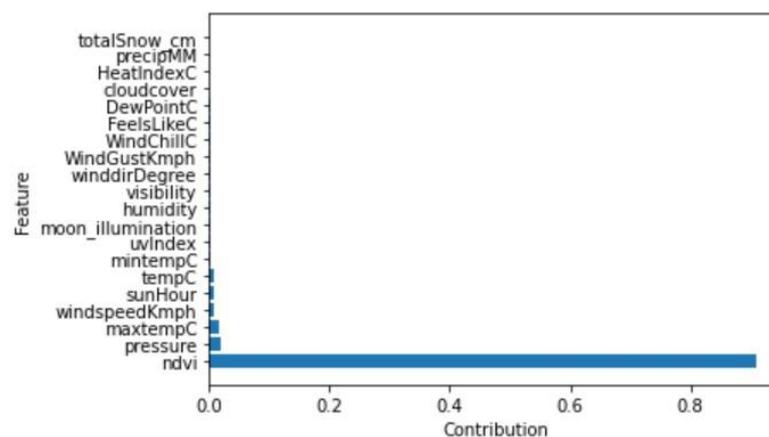


Рис. 2. – Гистограмма распределения важности отобранных признаков – NDVI прошлого периода и метеорологических параметров.

Из представленного графика видно, что значения NDVI за прошлый период в большей степени влияют на конечный результат. Рассмотрим на

рисунке 3 сравнение важности оставшихся метеорологических параметров на конечный результат.

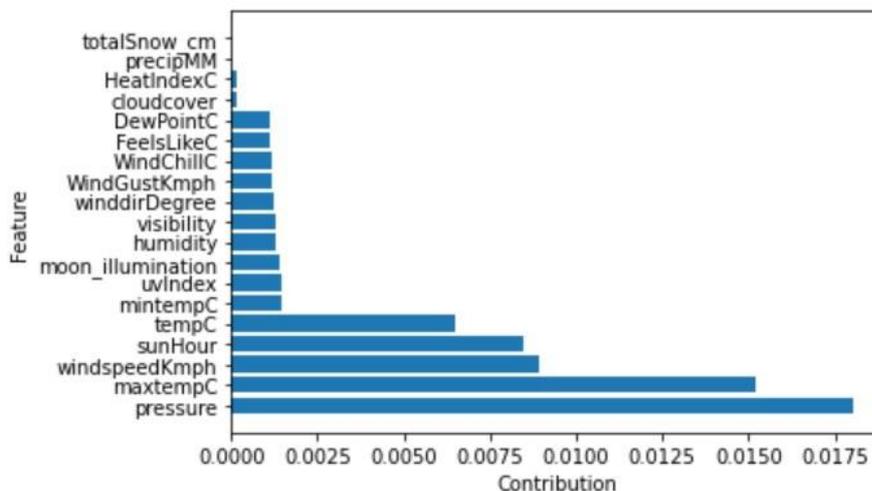


Рис. 3. – Гистограмма распределения важности метеорологических параметров.

Из представленного графика видно, что среди погодных признаков наиболее значимыми себя показали: давление, максимальная температура за сутки, средняя температура за сутки, скорость ветра и длительность солнечного дня.

Для оценки эффективности обученной модели применяются различные метрики. В данной работе используется метрика Root Mean Square Error (Среднеквадратичное отклонение, далее RMSE). Формула расчета RMSE представлена ниже:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

В качестве эталонных данных при расчете NDVI использовались вегетационные изображения из описанного выше датасета. В результате, средняя RMSE ошибка на тренировочных данных составила 0.20, а на валидационных данных – 0.27.

Чтобы учесть последовательно изменяющееся состояние растительности было принято решение рассмотреть в качестве модели для прогнозирования NDVI также рекуррентные нейронные сети с механизмом LSTM. Нейронные сети с подобной архитектурой также применялись в подобных исследованиях прогнозирования временных рядов [17], и конкретно временных рядов значений NDVI [13-14].

На рисунке 4 представлен график зависимости уменьшения ошибки RMSE от количества эпох в рамках лучшей попытки обучения модели LSTM.

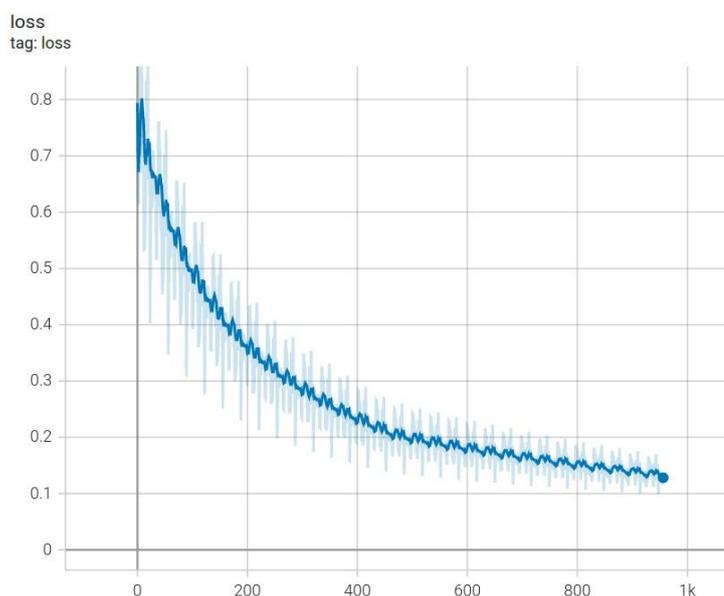


Рис. 4. – График зависимости RMSE от количества эпох обучения.

Ошибка RMSE при обучении модели LSTM составила 0.12 на тренировочной выборке, 0.14 – на валидационной выборке.

Таким образом, можно сделать вывод, что для прогнозирования NDVI и построения на его основе вегетационных карт лучше всего подходит нейросетевой метод прогнозирования, а именно рекуррентные нейронные сети с архитектурой LSTM. Результат работы в виде обученной модели планируется использовать в web-приложении для формирования вегетационных изображений будущего состояния растительности.

Литература

1. Костаринов, А. С., Латышенко Н. М. Точное земледелие. Наука молодых - будущее России : сборник научных статей 6-й Международной научной конференции перспективных разработок молодых ученых. Курск. Юго-Западный государственный университет. С. 341–344.

2. Гермак О.В. Использование данных дистанционного зондирования для экологического мониторинга опустынивания // Инженерный вестник Дона, 2013. №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2167.

3. Henrich, V., Krauss, G., Götze, C., Sandow, C. Entwicklung einer Datenbank für Fernerkundungsindize. Bochum. AK Fernerkundung. 2012. № 32. URL: indexdatabase.de/.

4. Huang, S., Tang, L., Hupy, J.P., Wang Y., Shao G. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. Journal of Forestry Research. 2021. URL: doi.org/10.1007/s11676-020-01155-1.

5. Xue J., Su B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. Journal of Sensors. 2017. URL: doi.org/10.1155/2017/1353691.

6. Катаев М. Ю., Бекеров А. А., Шалда П. В. Анализ трендов временных рядов вегетационного индекса NDVI. Доклады ТУСУР. 2017. №1. URL: cyberleninka.ru/article/n/analiz-trendov-vremennyh-ryadov-vegetatsionnogo-indeksa-ndvi.

7. Li S., Xu L., Jing Y., Yin H., Li X., Guan X. High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques. International Journal of Applied Earth Observation and Geoinformation. 2021. URL: doi.org/10.1016/j.jag.2021.102640.

8. Choudhary K., Shi W., Boori M. S., Corgne S. Agriculture Phenology Monitoring Using NDVI Time Series Based on Remote Sensing Satellites: A Case

Study of Guangdong, China. Optical Memory and Neural Networks. 2019. URL: doi.org/10.3103/S1060992X19030093.

9. Газизов Д. И. Обзор методов статистического анализа временных рядов и проблемы, возникающие при анализе нестационарных временных рядов // Научный журнал. 2016. №3. URL: cyberleninka.ru/article/n/obzor-metodov-statisticheskogo-analiza-vremennyh-ryadov-i-problemy-voznikayuschie-pri-analize-nestatsionarnyh-vremennyh-ryadov.

10. Fernández-Manso, A., Quintano, C., Fernández-Manso, O. Forecast of NDVI in coniferous areas using temporal ARIMA analysis and climatic data at a regional scale. *International Journal of Remote Sensing*. 2011. №32. Pp. 1595–1617.

11. Carreño-Conde, F., Sipols, A.E., de Blas, C.S., Mostaza-Colado, D. A forecast model applied to monitor crops dynamics using vegetation indices (Ndv). *Applied Sciences*. 2021. № 11.

12. Hachicha M., Louati M., Kallel A. International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). 2018. Pp. 1–4, URL: doi.org/10.1109/ATSIP.2018.8364499.

13. Cui C., Zhang W., Hong Z., Meng L. Forecasting NDVI in multiple complex areas using neural network techniques combined feature engineering. *International Journal of Digital Earth*. 2020. №13. URL: doi.org/10.1080/17538947.2020.1808718.

14. Cavalli, S., Penzotti, G., Amoretti, M., Caselli, S. A Machine Learning Approach for NDVI Forecasting based on Sentinel-2 Data. In *Proceedings of the 16th International Conference on Software Technologies – ICSOFT*. 2021. Pp. 473–480. URL: doi.org/10.5220/0010544504730480.

15. Li, X., Yuan, W., Dong, W. A Machine Learning Method for Predicting Vegetation Indices in China. *Remote Sensing*. 2021. №13. URL: doi.org/10.3390/rs13061147.

16. Чистяков С.П. Случайные леса: обзор. Труды Карельского научного центра РАН. 2013. №1. С. 117–136.
17. Передриенко, А.И., Лютая Т.М, Харитонов И.М, Степанченко И.В. Методы краткосрочного прогнозирования финансовых временных рядов с малыми объёмами выборки // Инженерный вестник Дона, 2020, №5. URL: ivdon.ru/ru/magazine/archive/N5y2020/6482.

References

1. Kostarinov, A. S., Latyshenok N. M. Tochnoe zemledelie. Nauka molodyh - budushchee Rossii: sbornik nauchnyh statej 6-j Mezhdunarodnoj nauchnoj konferencii perspektivnyh razrabotok molodyh uchenyh. Kursk. YUgo-Zapadnyj gosudarstvennyj universitet. pp. 341–344.
 2. Germak O.V. p Inzhenernyj vestnik Dona, 2013. №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2167.
 3. Henrich, V., Krauss, G., Götze, C., Sandow, C.. Bochum. AK Fernerkundung. 2012. № 32. URL: indexdatabase.de/.
 4. Huang, S., Tang, L., Hupy, J.P., Wang Y., Shao G. Journal of Forestry Research. 2021. URL: doi.org/10.1007/s11676-020-01155-1.
 5. Xue J., Su B. Journal of Sensors. 2017. URL: doi.org/10.1155/2017/1353691.
 6. Kataev M. YU., Bekerov A. A., SHalda P. V. Analiz trendov vremennyh ryadov vegetacionnogo indeksa NDVI. Doklady TUSUR. 2017. №1. URL: cyberleninka.ru/article/n/analiz-trendov-vremennyh-ryadov-vegetatsionnogo-indeksa-ndvi.
 7. Li S., Xu L., Jing Y., Yin H., Li X., Guan X. International Journal of Applied Earth Observation and Geoinformation. 2021. URL: doi.org/10.1016/j.jag.2021.102640.
-

8. Choudhary K., Shi W., Boori M. S., Corgne S. Agriculture Phenology Monitoring Using NDVI Time Series Based on Remote Sensing Satellites: A Case Study of Guangdong, China. *Optical Memory and Neural Networks*. 2019. URL: doi.org/10.3103/S1060992X19030093.
 9. Gazizov D. I. *Nauchnyj zhurnal*. 2016. №3. URL: cyberleninka.ru/article/n/obzor-metodov-statisticheskogo-analiza-vremennyh-ryadov-i-problemy-voznikayuschie-pri-analize-nestatsionarnyh-vremennyh-ryadov.
 10. Fernández-Manso, A., Quintano, C., Fernández-Manso, O. *International Journal of Remote Sensing*. 2011. №32. Pp. 1595–1617.
 11. Carreño-Conde, F., Sipols, A.E., de Blas, C.S., Mostaza-Colado, D. *Applied Sciences*. 2021. № 11.
 12. Hachicha M., Louati M., Kallel A. *International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. 2018. Pp. 1–4, URL: doi.org/10.1109/ATSIP.2018.8364499.
 13. Cui C., Zhang W., Hong Z., Meng L. *International Journal of Digital Earth*. 2020. №13. URL: doi.org/10.1080/17538947.2020.1808718.
 14. Cavalli, S., Penzotti, G., Amoretti, M., Caselli, S. A Machine Learning Approach for NDVI Forecasting based on Sentinel-2 Data. In *Proceedings of the 16th International Conference on Software Technologies. ICSOFT*. 2021. Pp. 473–480. URL: doi.org/10.5220/0010544504730480.
 15. Li, X., Yuan, W., Dong, W. A Machine Learning Method for Predicting Vegetation Indices in China. *Remote Sensing*. 2021. №13. URL: doi.org/10.3390/rs13061147.
 16. CHistyakov C.P. *Cluchajnye lesa: obzor. Trudy Karel'skogo nauchnogo centra RAN*. 2013. №1. pp. 117–136.
-



17. Peredrienko, A.I., Lyutaya T.M, Haritonov I.M, Stepanchenko I.V.
Inzhenernyj vestnik Dona, 2020, №5. URL:
ivdon.ru/ru/magazine/archive/N5y2020/6482.